(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 February 2005 (10.02.2005)

PCT

(10) International Publication Number
WO 2005/011474 A2

(51) International Patent Classification⁷: A61B

(21) International Application Number: PCT/US2004/024413

(22) International Filing Date: 30 July 2004 (30.07.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/491,524 1 August 2003 (01.08.2003) US
10/902,427 30 July 2004 (30.07.2004) US

(71) Applicants (for all designated States except US): CORRELOGIC SYSTEMS, INC. [US/US]; 6701 Democracy Boulevard, Suite 300, Bethesda, MD 20817 (US). THE UNITED STATES OF AMERICA as represented by the secretary of THE DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; c/o National Institute of Health, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): HITT, Ben, A. [US/US]; 929 Clopper Road, #B2, Gaithersburg, MD 20878 (US). LEVINE, Peter, A. [US/US]; 9608 Sotweed Dr., Potomac, MD 20854 (US). LIOTTA, Lance, A. [US/US]; 8601 Bradley Blvd., Bethesda, MD 20817 (US). PETRICOIN, Emanuel, F. [US/US]; 2805 Feather Ridge Ct., Dunkirk, MD 20754 (US).

(74) Agent: FORD, Timothy, D.; c/o Cooley Godward, LLC, One Freedom Square, Reston Town Center, 11951 Freedom Drive, Reston, VA 20190-5656 (US).

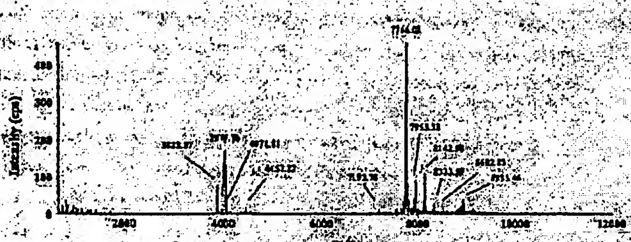
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: MULTIPLE HIGH-RESOLUTION SERUM PROTEOMIC FEATURES FOR OVARIAN CANCER DETECTION



(57) Abstract: A well-controlled serum study set (n = 248) from women being followed and evaluated for the presence of ovarian cancer was used to extend serum proteomic pattern analysis to a higher resolution mass spectrometer instrument platform to explore the existence of multiple distinct highly accurate diagnostic sets of features present in the same mass spectrum. Multiple highly accurate diagnostic proteomic feature sets exist within human sera mass spectra. Using high-resolution mass spectral data, at least 56 different patterns were discovered that achieve greater than 85 % sensitivity and specificity in testing and validation. Four of those feature sets exhibited 100 % sensitivity and specificity in blinded validation. The sensitivity and specificity of diagnostic models generated from high-resolution mass spectral data were superior ($P < 0.00001$) than those generated from low-resolution mass spectral data using the same input sample.

WO 2005/011474 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

BEST AVAILABLE COPY

Multiple High-resolution Serum Proteomic Features for Ovarian Cancer Detection

Background

[1001] Serum proteomic pattern analysis by mass spectrometry (MS) is an emerging technology that is being used to identify biomarker disease profiles. Using this MS-based approach, the mass spectra generated from a training set of serum samples is analyzed by a bioinformatic algorithm to identify diagnostic signature patterns comprised of a subset of key mass-to-charge (*m/z*) species and their relative intensities. Mass spectra from unknown samples are subsequently classified by likeness to the pattern found in mass spectra used in the training set. The number of key *m/z* species whose combined relative intensities define the pattern represent a very small subset of the entire number of species present in any given serum mass spectrum.

[1002] The feasibility of using MS proteomic pattern analysis for the diagnosis of ovarian, breast, and prostate cancer has been demonstrated. While investigators have used a variety of different bioinformatic algorithms for pattern discovery, the most common analytical platform is comprised of a low-resolution time-of-flight (TOF) mass spectrometer where samples are ionized by surface enhanced laser desorption/ionization (SELDI), a ProteinChip array-based chromatographic retention technology that allows for direct mass spectrometric analysis of analytes retained on the array.

[1003] Ovarian cancer is the leading cause of gynecological malignancy and is the fifth most common cause of cancer-related death in women. The American Cancer Society estimates that that there will be 23,300 new cases of ovarian cancer and 13,900 deaths in 2002. Unfortunately, almost 80% of women with common epithelial ovarian cancer are not diagnosed until the disease is advanced in stage, i.e., has spread to the upper abdomen (stage III) or beyond (stage IV). The 5-year survival rate for these women is only 15 to 20%, whereas the 5-year survival rate for ovarian cancer at stage I approaches 95% with surgical intervention. The early diagnosis of ovarian cancer, therefore, could dramatically decrease the number of deaths from this cancer.

[1004] The most widely used diagnostic biomarker for ovarian cancer is Cancer Antigen 125 (CA 125) as detected by the monoclonal antibody OC 125. Though 80% of patients with ovarian cancer possess elevated levels of CA 125, it is elevated in only 50-60% of patients at stage I, lending it a positive-predictive value of 10%. Moreover, CA 125 can be elevated in other non-gynecologic and benign conditions. A combined strategy of CA 125 determination with ultrasonography increases the positive-predictive value to approximately 20%.

[1005] Low molecular weight serum proteomic patterns from low-resolution SELDI-TOF MS data can distinguish neoplastic from non-neoplastic disease within the ovary. See Petricoin, E. F. III *et al.* Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 359, 572-577 (2002). The proteomic patterns can be identified by application of an artificial intelligence bioinformatics tool that employs an unsupervised system (self-organizing cluster mapping) as a fitness test for a supervised system (a genetic algorithm). A training set comprised of SELDI-TOF mass spectra from serum derived from either unaffected women or women with ovarian cancer is employed so that the most fit combination of *m/z* features (along with their relative intensities) plotted in n-space can reliably distinguish the cohorts used in training. The "trained" algorithm is applied to a masked set of samples that resulted in a sensitivity of 100% and a specificity of 95%. This technique is described in more detail in WO 02/06829A2 "A Process for Discriminating Between Biological States Based on Hidden Patterns From Biological Data" ("Hidden Patterns") the disclosure of which is hereby expressly incorporated herein by reference.

[1006] Although this technique works well, the low-resolution mass spectrometric instrumentation and thus the data that comes from the instrument may limit the attainable reproducibility, sensitivity, and specificity for proteomic pattern analyses for routine clinical use.

Summary

[1007] The protein pattern analysis concept of Hidden Patterns is extended to a high-resolution MS platform to generate diagnostic models possessing higher sensitivities and

specificities on a format that generates more stable spectra, has a true time-of-flight mass accuracy, and is inherently more reproducible machine-to-machine and day-to-day because of the increase in mass accuracy. Sera from a large, well-controlled ovarian cancer screening trial were used and proteomic pattern analysis was conducted on the same samples on two mass spectral platforms differing in their effective resolution and mass accuracy. The data was analyzed so as to rank the sensitivity and specificity of the series of diagnostic models that emerged.

[1008] The spectra from a high-resolution and a low-resolution mass spectrometer with the same patients' sera samples applied and analyzed on the same SELDI ProteinChip arrays were compared. Although the higher resolution mass spectra may generate more distinguishable sets of diagnostic features, the increased complexity and dimensionality of data may reduce the likelihood of fruitful pattern discovery. Diagnostic proteomic feature sets can be discerned within the high-resolution spectra from the clinically relevant patient study set, and the modeling outcomes between the two instrument platforms can be compared. The number and character of the diagnostic models emerging from data mining operations can be ranked. Serum proteomic pattern analysis can be used for the generation of multiple, highly accurate models using a hybrid quadrupole time-of-flight (Qq-TOF) MS for an improved early diagnosis of ovarian cancer.

Brief Description of the Figures

[1009] FIGS. 1A and 1B compare the mass spectra from control serum prepared on a WCX2 ProteinChip array and analyzed with a PBS-II TOF (panel A) or a Qq-TOF (panel B) mass spectrometer.

[1010] FIGS. 2A and 2B show histograms representing the testing results of sensitivity (2A) and specificity (2B) of 108 models for MS data acquired on either a Qq-TOF or a PBS-II TOF mass spectrometer.

[1011] FIGS. 3A and 3B show histograms representing the testing and blinded validation results of sensitivity (3A) and specificity (3B) of 108 models for MS data acquired on either a Qq-TOF or a PBS-II TOF mass spectrometer.

[1012] FIGS. 4A and 4B compare SELDI Qq-TOF mass spectra of serum from an unaffected individual (4A) and an ovarian cancer patient (4B).

Detailed Description

Analysis of Serum Samples

[1013] A total of 248 serum samples were provided from the National Ovarian Cancer Early Detection Program (NOCEDP) clinic at Northwestern University Hospital (Chicago, Illinios). The samples were processed and their proteomic patterns acquired by MS as described below in the description of the methods used. The serum samples in the present study were analyzed on the same protein chip arrays by both a PBS-II and a Qq-TOF MS fitted with a SELDI ProteinChip array interface. While the spectra acquired from both instruments are qualitatively similar, the higher resolution afforded by the Qq-TOF MS is apparent from FIG. 1. This increased resolution allows species close in *m/z* unresolved by the PBS-II TOF MS to be distinctly observed in the Qq-TOF mass spectrum. Indeed, simulations demonstrate the ability of the Qq-TOF MS (routine resolution ~ 8000) to completely resolve species differing in *m/z* of only 0.375 (i.e., at *m/z* 3000) whereas complete resolution of species with the PBS-II TOF MS (routine resolution ~ 150) is only possible for species that differ by *m/z* of 20 (simulation not shown).

[1014] The mass spectra were analyzed using the ProteomeQuest™ bioinformatics tool employing ASCII files consisting of *m/z* and intensity values of either the PBS-II TOF or the Qq-TOF mass spectra as the input. The mass spectral data acquired using the Qq-TOF MS were binned to precisely define the number of features in each spectrum to 7,084 with each feature being comprised of a binned *m/z* and amplitude value. The algorithm examines the data to find a set of features at precise binned *m/z* values whose combined, normalized relative intensity values in n-space best segregate the data derived

from the training set. Mass spectra acquired on the Qq-TOF and the PBS-II TOF instruments from the same sample sets were restricted to the *m/z* range from 700 to 11,893 for direct comparison between the two platforms. The entire set of spectra acquired from the serum samples was divided into three data sets: a) a training set that is used to discover the hidden diagnostics patterns, b) a testing set, and c) a validation set. With this approach only the normalized intensities of the key subset of *m/z* values identified using the training set were used to classify the testing and validation sets, and the algorithm had not previously "seen" the spectra in the testing and validation sets.

[1015] The training set was comprised of serum from 28 unaffected women and 56 women with ovarian cancer. The training and testing set mass spectra were analyzed by the bioinformatic algorithm to generate a series of models under the following set modeling parameters: a) a similarity space of 85%, 90%, or 95% likeness for cluster classification; b) a feature set size of 5, 10, or 15 random *m/z* values whose combined intensities comprise each pattern; and c) a learning rate of 0.1%, 0.2%, or 0.3% for pattern generation by the genetic algorithm. Four sets of randomly generated models for each of the 27 permutations were derived and queried with the same test set. Sensitivity and specificity testing results for each of the 108 models (four rounds of training for each of the 27 permutations) were generated, as shown in FIGS. 2A and 2B. These results demonstrate that the Qq-TOF MS data produced better results than the lower resolution spectra ($P < 0.00001$, using the exact Cochran-Armitage test (see Agresti A. Categorical Data Analysis New York: John Wiley and Sons (1990)) for trend) throughout a range of modeling conditions.

[1016] The ability to generate the best performing models for testing and validation was statistically evaluated as multiple models were generated and ranked using the entire range of the modeling parameters above. Models from the training set were validated using a testing set consisting of 31 unaffected and 63 ovarian cancer serum samples. To further validate the ability to diagnose ovarian cancer, a set of blinded sample mass spectra consisting of an additional 37 normal and 40 ovarian cancer serum mass spectra were tested against the model found in training previously discussed. As shown in FIGS. 3A and 3B, the results show the ability of the mass spectra from the higher

resolution Qq-TOF MS to generate statistically significant ($P < 0.00001$) superior models over the lower resolution PBS-II mass spectra.

[1017] Fifteen models were found that were 100% sensitive in their ability to correctly discriminate unaffected women from those suffering from ovarian cancer, that were 100% specific in discriminating women in the test set, and at least 97% specific in the validation set. These models are shown in Appendix A, and identified as Model 1 through Model 15. Of these models, four were found that were both 100% sensitive and specific for both sets (Models 4, 9, 10, and 15).

[1018] Appendix A identifies for each model the following information: First the specificity and sensitivity for each model is shown for the Test set and for the Validity set. The number of samples for which the model correctly grouped women with a "Normal State" (i.e. not having ovarian cancer) and with an "Ovarian Cancer State" is then shown for each of the test and validity tests, compared to the total number of samples in the corresponding sets. For example, in Model 1, the model correctly identified 36 of the 37 women as having a normal state in the Validity set.

[1019] Finally, for each model a table is set forth showing the constituent "patterns" comprising the model. Each pattern corresponds to a point, or node, in the N-dimensional space defined by the N *m/z* values (or "features") included in the model. Thus, each pattern is a set of features, each feature having an amplitude. Appendix A therefore shows for each model a table containing the constituent patterns, each pattern being in a row identified by a "Node" number. The table also includes columns for the constituent features of the patterns, with the *m/z* value for each pattern identified at the top of the column. The amplitudes are shown for each feature, for each pattern, and are normalized to 1.0. The remaining four columns in each table are labeled "Count," "State," "StateSum," and "Error." "Count" is the number of samples in the Training set that correspond to the identified node. "State" indicates the state of the node, where 1 indicates diseased (in this case, having ovarian cancer) and 0 indicates normal (not having the disease). "StateSum" is the sum of the state values for all of the correctly classified members of the indicated node, while "Error" is the number of incorrectly

classified members of the indicated node. Thus, for node 5 in Model 1, 13 samples were assigned to the node, whereas 11 samples were actually diseased. StateSum is thus 11 (rather than 13) and Error is 2.

[1020] Examination of the key *m/z* features that comprise the four best performing models (Models 4, 9, 10, and 15) reveals certain features (i.e., contained within *m/z* bins 7060.121, 8605.678 and 8706.065) that are consistently present as classifiers in those models.

[1021] Although the proteomic patterns generated from both healthy and cancer patients using the Qq-TOF MS are quite similar (as seen by comparing FIGS. 4A to 4B), careful inspection of the raw mass spectra reveals that peaks within the binned *m/z* values 7060.121 and 8605.678 are differentially abundant in a selection of the serum samples obtained from ovarian cancer patients as compared to unaffected individuals and that the features that the ProteomeQuest™ software selected are "real" features and not noise. The insets in FIGS. 4A and 4B show expanded *m/z* regions highlighting significant intensity differences of the peaks in the *m/z* bins 7060.121 and 8605.678 (indicated by brackets) identified by the algorithm as belonging to the optimum discriminatory pattern. These results indicate these MS peaks originate from species that may be consistent indicators of the presence of ovarian cancer. The ability to distinguish sera from an unaffected individual or an individual with ovarian cancer based on a single serum proteomic *m/z* feature alone, however, is not possible across the entire serum study set. While a single key *m/z* species is insufficient to globally distinguish all of the unaffected and ovarian cancer patients, taken together the combined peak intensities of key ions does allow the two data sets to be completely distinguished.

[1022] The four best performing models that are 100% sensitive and specific for the blinded testing and validation tests were chosen for further analysis. Table 1 shows bioinformatic classification results of serum samples from masked testing and validation sets by proteomic pattern classification using the best performing models.

	Actual	Predicted (%)
Benign / Unaffected	66	66 (100)
Ovarian Cancer Stage I	22	22 (100)
Ovarian Cancer Stage II, III, IV	81	81 (100)

Table 1

Each of these models was able to successfully diagnose the presence of ovarian cancer in all of the serum samples from affected women. Further, no false positive or false negative classifications occurred with these best performing models.

Discussion

[1023] A limitation of individual cancer biomarkers is the lack of sensitivity and specificity when applied to large heterogeneous populations. Biomarker pattern analysis seeks to overcome the limitation of individual biomarkers. Serum proteomic pattern analysis can provide new tools for early diagnosis, therapeutic monitoring and outcome analysis. Its usefulness is enhanced by the ability of a selected set of features to transcend the biologic heterogeneity and methodological background "noise." This diagnostic goal is aided by employing a genetic algorithm coupled with a self-organizing cluster analysis to discover diagnostic subsets of *m/z* features and their relative intensities contained within high-resolution Qq-TOF mass spectral data.

[1024] It is believed that diagnostic serum proteomic feature sets exist within constellations of small proteins and peptides. A given signature pattern reflects changes in the physiologic or pathologic state of a target tissue. With regard to cancer markers, it is believed that serum diagnostic patterns are a product of the complex tumor-host microenvironment. It is thought likely that the set of diagnostic features is partially derived from multiple modified host proteins rather than emanating exclusively from the cancer cells. The biomarker profile may be amplified by tumor-host interactions. This amplification includes, for example, the generation of peptide cleavage products by tumor or host proteases. There may exist multiple dependent or independent sets of proteins/peptides that reflect the underlying tissue pathology. Hence, the disease related proteomic pattern information content in blood might be richer than previously anticipated. Rather than a single "best" feature set, multiple proteomic feature sets may exist that achieve highly accurate discrimination and hence diagnostic power. This possibility is supported by the data described above.

[1025] The low molecular weight serum proteome is an unexplored archive, even though this is the mass region where MS is best suited for analysis. It is thought likely that disease-associated species are comprised of low molecular weight peptide/protein species that vary in mass by as little as a few Daltons. Thus a higher resolution mass spectrometer would be expected to discriminate and discover patterns not resolvable by a lower resolution instrument. The spectra produced by a Qq-TOF MS were compared to that of the Ciphergen PBS-II TOF MS. The routine resolution obtained is in excess of 8000 (at m/z = 1500) for the Qq-TOF MS and 150 (at m/z = 1500) for the PBS-II TOF mass spectrometer. A SELDI source was used so that both instruments analyzed the same sample on distinct regions of the protein chip array bait surface. While the overall spectral profile is similar, a single peak on the PBS-II TOF MS is resolved into a multitude of peaks on the Qq-TOF MS (seen by comparing FIGS 1A and 1B to FIGS. 4A and 4B). Moreover, the inherent increase in mass accuracy by higher resolution instrumentation that has uncoupled the mass analyzer from the source will provide for cleaner spectra as this will suppress confounding, metastable ions, generate spectra with lower mass drift over time and instruments at the same time as generating more complex, highly resolved data.

[1026] In the first phase of comparison, proteomic patterns from mass spectra derived from the same training sets and generated on the high and low resolution mass spectrometers were scrutinized for their overall sensitivity and specificity over a series of modeling constraints in which patterns were generated using three different degrees of similarity space for the self-organizing clusters to form, three different sets of feature sizes chosen, and three different mutation rates for a total of 27 modeling permutations. Sensitivity and specificity testing results for each of the 108 models (shown in FIGS. 2A and 2B), produced from four rounds of training for each of the 27 permutations, demonstrate that the Qq-TOF MS generated spectra consistently outperformed the lower resolution TOF-MS spectra ($P < 0.00001$) independent of the modeling criteria used.

[1027] Since the spectra from the higher resolution platform generate patterns with a higher level of sensitivity and specificity, those spectra could generate more accurate models with a higher degree of sensitivity and specificity - that is, generate the best

diagnostic models. These results were generated using even more stringent criteria, in that an additional masked validation set was employed after testing to determine overall accuracy. The higher resolution spectra consistently produced significantly more accurate models as seen in both the testing and validation studies (as shown in FIGS. 3A and 3B). The models derived from the Qq-TOF MS were consistently more sensitive and specific ($P < 0.00001$) than those from the PBS-II TOF MS. Four models were generated that attained 100% sensitivity and specificity in both testing and validation. The number of key m/z values used as classifiers in the four best diagnostic models ranged from 5 to 9. Three m/z bin values were found in two of these four models and two m/z bins were found in three of the four best models. The distinct peaks present in the recurring m/z bins 7060.121, 8605.678 and 8706.065 may be good candidates for low molecular weight components in serum that may be key disease progression indicators.

[1028] These data support the existence of multiple highly accurate and distinct proteomic feature sets that can accurately distinguish ovarian cancer. To screen for diseases of relatively low prevalence, such as ovarian cancer, a diagnostic test preferably exceeds 99% sensitivity and specificity to minimize false positives, while correctly detecting early stage disease when it is present. As discussed above, four models generated using high-resolution Qq-TOF MS data achieved 100% sensitivity and specificity. In blinded testing and validation studies any one of these models were used to correctly classify 22/22 stage-I ovarian cancer, 81/81 ovarian cancer stage II, III and IV and 68/68 benign disease controls.

[1029] Thus, a clinical test could simultaneously employ several combinations of highly accurate diagnostic proteomic patterns arising concomitantly from the same data streams, which, taken together, could achieve an even higher degree of accuracy in a screening setting where a diagnostic test will face large population heterogeneity and potential variability in sample quality and handling. Hence, a high-resolution system, such as the Qq-TOF MS employed in this study, is preferred based on the present results.

Methods

[1030] *Serum Samples*: Serum samples were obtained from the National Ovarian Cancer Early Detection Program (NOCEDP) clinic at Northwestern University Hospital (Chicago, Illinois). Two hundred and forty eight samples were prepared using a Biomek 2000 robotic liquid handler (Beckman Coulter, Inc., Palo Alto, California). All analyses were performed using ProteinChip weak cation exchange interaction chips (WCX2, Ciphergen Biosystems Inc., Fremont, California). A control sample was randomly applied to one spot on each protein array as a quality control for sample preparation and mass spectrometer function. The control sample, SRM 1951A, which is comprised of pooled human sera, was provided by the National Institute of Standards and Technology (NIST).

[1031] *Sample Preparation*: WCX2 ProteinChip arrays were processed in parallel using a Biomek Laboratory workstation (Beckman-Coulter) modified to make use of a ProteinChip array bioprocessor (Ciphergen Biosystems Inc.). The bioprocessor holds 12 ProteinChips, each having 8 chromatographic "spots", allowing 96 samples to be processed in parallel. One hundred µl of 10 mM HCL was applied to the WCX2 protein arrays and allowed to incubate for 5 minutes. The HCl was aspirated, discarded and 100 µl of distilled, deionized water (ddH_2O) was applied and allowed to incubate for 1 minute. The ddH_2O was aspirated, discarded, and reapplied for another minute. One hundred µl of 10 mM NH_4HCO_3 with 0.1% Triton X-100 was applied to the surface and allowed to incubate for 5 minutes after which the solution was aspirated and discarded. A second application of 100 µL of 10 mM NH_4HCO_3 with 0.1% Triton X-100 was applied and allowed to incubate for 5 minutes after which the ProteinChip array bait surfaces were aspirated. Five µl of raw, undiluted serum was applied to each ProteinChip WCX2 bait surface and allowed to incubate for 55 minutes. Each ProteinChip array was washed 3 times with Dulbecco's phosphate buffered saline (PBS) and ddH_2O. For each wash, 150 µl of either PBS or ddH_2O was sequentially dispensed, mixed by aspirating and dispensed for a total of 10 times in the bioprocessor after which the solution was aspirated to waste. This wash process was repeated for a total of 6 washes per ProteinChip array bait surface. The ProteinChip array bait surfaces were vacuum dried to prevent cross contamination when the bioprocessor gasket was removed. After removing

the bioprocessor gasket, 1.0 µl of a saturated solution of α-cyano-5-hydroxycinnamic acid in 50% (v/v) acetonitrile, 0.5% (v/v) trifluoroacetic acid was applied to each spot on the ProteinChip array twice, allowing the solution to dry between applications.

[1032] *PBS-II Analysis*: ProteinChip arrays were placed in the Protein Biological System II time-of-flight mass spectrometer (PBS-II, Ciphergen Biosystems Inc.) and mass spectra were recorded using the following settings: 195 laser shots/spectrum collected in positive mode, laser intensity 220, detector sensitivity 5, detector voltage 1850, and a mass focus of 6,000 Da. The PBS-II was externally calibrated using the "All-In-One" peptide mass standard (Ciphergen Biosystems, Inc.).

[1033] *Qq-TOF MS Analysis*: ProteinChip arrays were analyzed using a hybrid quadrupole time-of-flight mass spectrometer (QSTAR pulsar I, Applied Biosystems Inc., Framingham, Massachusetts) fitted with a ProteinChip array interface (Ciphergen Biosystems Inc., Fremont, California). Samples were ionized with a 337 nm pulsed nitrogen laser (ThermoLaser Sciences model VSL-337-ND-S, Waltham, Massachusetts) operating at 30 Hz. Approximately 20 mTorr of nitrogen gas was used for collisional ion cooling. Each spectrum represents 100 multi-channel averaged scans (1.667 min acquisition/spectrum). The mass spectrometer was externally calibrated using a mixture of known peptides.

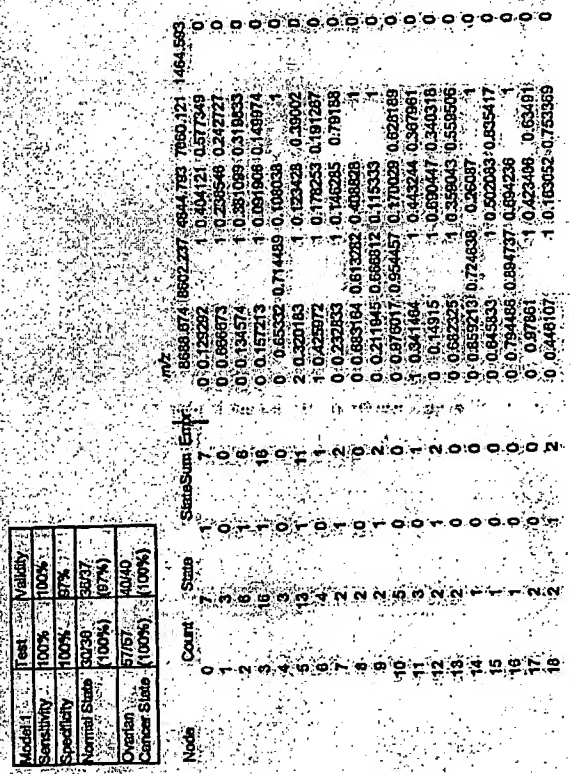
[1034] *Proteomic Pattern Analysis*: Proteomic pattern analysis was performed by exporting the raw data file generated from the Qq-TOF mass spectrum into a tab-delimited format that generated approximately 350,000 data points per spectrum. The data files were binned using a function of 400 parts per million (ppm) such that all data files possess identical *m/z* values (e.g., the *m/z* bin sizes linearly increased from 0.28 at *m/z* 700 to 4.75 at *m/z* 12,000). The intensities in each 400 ppm bin were summed. This binning process condenses the number of data points to exactly 7,084 points per sample. The binned spectral data were separated into approximately three equal groups for training, testing and blind validation. The training set consisted of 28 normal and 56 ovarian cancer samples. The models were built on the training set using ProteomeQuest™ (Correlogic Systems Inc., Bethesda, Maryland) and validated using the

testing samples, which consisted of 30 normal and 57 ovarian cancer samples. The model was validated using blinded samples, which consisted of 37 normal and 40 ovarian cancer samples. These *m/z* values that were found to be classifiers used to distinguish serum from a patient with ovarian cancer from that of an unaffected individual are based on the binned data and not the actual *m/z* values from the raw mass spectra.

[1035] Statistical significance of the results generated using the Qq-TOF and PBS-II MS was performed using the exact Cochran-Armitage test for trend to compare the distributions of these specificity and sensitivity values between the two instrumental platforms evaluated since the models are constructed independently from each other.

Appendix A

Model1	Test	Validity
Sensitivity	100%	100%
Specificity	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)



Model 2	Test:	Validity:
Sensitivity:	100%	100%
Specificity:	100%	95%
Normal State	30/30 (100%)	35/37 (95%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 8605.678	5773.642	6256.91	7060.121	8706.065	748.048
0	7	1	7	0	0.836245	0.103496	0.112529	0.966826	0.445348	0
1	3	0	0	0	0.991918	0.304599	0.273147	0.468784	0.965068	0
2	10	1	10	0	0	1	0.069882	0.103221	0.545584	0.405998	0
3	3	0	0	0	0.668897	0.155636	0.241726	0.965206	0.964241	0
4	13	1	8	5	0.966501	0.107261	0.192038	0.625891	0.857142	0
5	3	1	3	0	0.595203	0.103657	0.125339	1	0.430678	0
6	2	0	0	0	0.610908	0.26603	0.555267	0.974007	1	0
7	3	1	3	0	0.894877	0.117567	0.231772	1	0.818855	0
8	8	1	8	0	0	1	0.12112	0.122806	0.745443	0.523198	0
9	7	0	0	0	0.69096	0.178288	0.258833	0.503651	1	0
10	10	1	10	0	0	1	0.047377	0.061628	0.284495	0.406995	0
11	1	0	0	0	0	1	0.133102	0.208333	0.305556	0.803241	0
12	4	0	0	0	0.59657	0.159346	0.30219	0.707978	1	0
13	1	1	1	0	0.411765	0.12549	0.137255	1	0.266667	0
14	1	0	0	0	0.819951	0.311436	0.408759	1	0.961071	0
15	1	0	0	0	0.865909	0.315909	0.404545	0.711364	1	0

WO 2005011474
PCT/US2004023413

Model 3	Test:	Validity:
Sensitivity:	100%	100%
Specificity:	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 8605.678	6668.674	6345.106	9843.343	3354.185	7374.687	5968.506
0	30	1	29	1	1	0.387229	0.015495	0.128241	0.006888	0.213168	0.032564
1	6	0	1	1	1	0.813858	0.049105	0.268494	0.031062	0.354791	0.060409
2	5	0	0	0	0.943078	0.9957	0.023126	0.32079	0.05742	0.600263	0.033526
3	19	1	14	5	1	0.582078	0.049422	0.20029	0.026914	0.389413	0.026103
4	1	0	0	0	0.918669	1	0.042514	0.260628	0.170055	0.914972	0
5	1	0	0	0	0.820513	1	0.125356	0	0.333333	0.948718	0.321937
6	3	1	3	0	0	1	0.715204	0.006153	0.19098	0.060695	0.722323	0.025888
7	1	1	1	0	0	1	0.573192	0	0	0.151675	0.130511	0.982363	0.044092
8	3	0	0	0	0.937262	0.9936	0.115137	0.159158	0	0.830834	0.113328
9	3	0	0	0	0.722109	1	0.017883	0.045724	0.057432	0.817682	0.059098
10	1	0	0	0	0.950943	1	0.320755	0.230189	0	0.664151	0.301887
11	2	1	2	0	0	1	0.41404	0.079637	0.146901	0.038536	0.645357	0
12	1	0	0	0	0.980798	1	0.075332	0.51551	0	0.401773	0.025111
13	1	0	0	0	0.906907	1	0.081081	0.012012	0.189189	0.429429	0

WO 2005011474
PCT/US2004023413

Model 4	Test	Validity
Sensitivity	100%	100%
Specificity	100%	100%
Normal State	30/30 (100%)	37/37 (100%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 7060.121	7096.922	8605.676	6548.771	8706.065	818.4801	8540.536	8352.723
0	8	1	8	0	0.917113	0.21551	0.661398	0.121208	0.444445	0	0.518113	0.110812
1	3	0	0	0	0.492091	0.305348	0.968398	0.205158	0.894171	0	0.951383	0.236869
2	10	1	10	0	0.547569	0.173669	1	0.104231	0.409816	0	0.51695	0.092658
3	3	0	0	0	0.929844	0.33378	0.674226	0.166695	0.963815	0	0.90104	0.157423
4	8	1	8	0	0.732832	0.276296	1	0.135825	0.570368	0	0.883495	0.107333
5	10	1	7	3	0.648923	0.304081	0.983209	0.148318	0.82452	0	0.916506	0.12435
6	3	0	0	0	0.346591	0.221128	1	0.173951	0.806024	0	0.827509	0.179187
7	4	1	4	0	1	0.262028	0.56594	0.124258	0.40729	0	0.422331	0.10647
8	2	0	0	0	0.794377	0.531631	0.515983	0.290957	0.814304	0	1	0.29799
9	1	1	1	0	1	0.270158	0.532108	0.145686	0.831683	0	0.946252	0.132956
10	6	0	0	0	0.437313	0.281307	0.815518	0.170126	0.890092	0	0.886262	0.143115
11	10	1	10	0	0.282368	0.113517	1	0.06052	0.405555	0	0.507878	0.047164
12	3	0	0	0	0.652298	0.545467	0.758154	0.391447	0.993289	0	0.878634	0.361204
13	3	0	0	0	0.663094	0.35973	0.501834	0.214181	0.872978	0	1	0.191813
14	2	0	1	1	1	1	0.636476	0.845795	0.372277	0.937743	0	0.965217	0.311208
15	1	1	1	0	1	1	0.237154	0.735178	0.105402	0.753623	0	0.756258	0.102767

WO 2005/111474 PCT/US2005/014113

17

Model 5	Test	Validity
Sensitivity	100%	100%
Specificity	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 11601.83	8716.517	3419.205	4260.403	1229.752	2007.145	8602.237	7060.121	846.10
0	30	1	30	0	0.045973	0.188825	0.031338	0.084657	0.008804	0.010191	1	0.232181	0.0142
1	2	0	0	0	0.190458	0.752349	0.206444	0.438551	0	0.0639	1	0.321633	0.3785
2	2	0	0	0	0.195637	0.728544	0.15697	0.355302	0	0.029894	0.730036	1	0.052(
3	17	1	11	6	0.076996	0.33797	0.088986	0.20709	0.029195	0.022459	1	0.437262	0.043(
4	2	0	0	0	0.115091	0.512947	0.110247	0.353618	0.002048	0.043823	1	0.230496	0.209(
5	5	1	5	0	0.090591	0.287811	0.087215	0.154745	0.015448	0.049325	1	0.740332	0.014(
6	1	0	0	0	0.202229	0.542994	0.402866	0.52707	0.197452	0	0.621019	1	0.259(
7	2	1	2	0	0.106417	0.226812	0.165819	0.205581	0.014039	0.018811	0.69384	1	0.035(
8	2	0	0	0	0.143113	1	0.214746	0.826275	0.086968	0	0.92163	0.582288	0.483(
9	1	0	0	0	0.178571	0.921053	0.274436	0.744361	0	0.087669	1	0.772556	0.24(
10	2	0	0	0	0.127322	0.855385	0.288389	0.341074	0.000943	0.066154	0.973585	0.601901	0.555(
11	3	0	0	0	0.230129	0.726008	0.280667	0.633693	0.045805	0.024148	0.754434	1	0.104(
12	2	0	0	0	0.18007	0.762553	0.209338	0.57439	0	0.086841	1	0.675463	0.400(
13	1	0	0	0	0.127701	0.565815	0.125737	0.675835	0.037328	0	1	0.844794	0.149(
14	1	0	0	0	0.138095	0.784127	0.163492	0.477778	0	0.014286	1	0.760317	0.063(
15	1	0	0	0	0.291045	0.808458	0.271144	0.41791	0	0.014925	0.895522	1	0.383(
16	1	0	0	0	0.158163	0.785714	0.318878	0.558673	0	0.035714	1	0.612245	0.877(
17	2	1	2	0	0.154471	0.472129	0.131158	0.216488	0.027697	0	1	0.784209	0.167(

WO 2005/111474 PCT/US2005/014113

18

Model 8	Test	Validity
Sensitivity	100%	100%
Specificity	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 8688.674	8602.237	7060.121	4920.131	10431.02	2817.487
0	12	1	12	0	0.212098	1	0.44326	0.05893	0.243359	0
1	2	0	0	0	0.7195	1	0.320383	0.194066	0.325602	0
2	19	1	19	0	0.181351	1	0.188047	0.02488	0.074401	0
3	6	0	0	0	0.721687	0.728508	1	0.146458	0.244383	0
4	7	1	5	2	0.328961	1	0.392833	0.054395	0.118492	0
5	8	1	6	2	0.430797	1	0.446852	0.061423	0.263857	0
6	4	0	0	0	0.479383	1	0.241389	0.13775	0.164372	0
7	3	1	3	0	0.265618	1	0.781812	0.070789	0.199972	0
8	1	1	1	0	0.264706	0.703013	1	0.066715	0.351606	0
9	1	1	1	0	0.218579	1	0.872131	0.213115	0.464481	0
10	6	0	0	0	0.979239	0.960158	0.668689	0.134247	0.169243	0
11	2	0	0	0	0.687882	1	0.587495	0.248281	0.240037	0
12	1	1	1	0	0.195426	0.60499	1	0.04252	0.096674	0
13	1	0	0	0	0.686347	1	0.854244	0.156827	0.560885	0
14	1	0	0	0	0.788458	0.890625	1	0.330729	0.5625	0
15	1	0	0	0	0.987805	1	0.536585	0.140244	0	0
16	1	1	1	0	0.486765	1	0.741176	0.068177	0.448529	0
17	1	1	1	0	0.478368	1	0.888279	0.068999	0.25958	0

Model 8	Test	Validity
Sensitivity	100%	100%
Specificity	100%	95%
Normal State	30/30 (100%)	35/37 (95%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 708.1657	8605.678	6606.643	7060.121	6761.677	2472.108	8708.065	5511.917	1195.325	500
0	9	1	9	0	0	0.978759	0.129335	0.890026	0.141874	0.08438	0.465115	0.117064	0.112831	0.0
1	5	0	0	0	0	0.994064	0.168514	0.384269	0.247993	0.076075	0.898872	0.147354	0.126049	0.1
2	15	1	15	0	0	1	0.092694	0.597216	0.154853	0.061148	0.463791	0.081717	0.104318	0.0
3	4	0	0	0	0	0.660345	0.19312	0.967633	0.301109	0.102143	0.97033	0.184698	0.154734	0.1
4	12	1	8	4	0.000896	0.966228	0.160728	0.635568	0.230458	0.048255	0.860368	0.09372	0.147295	0.0
5	4	1	4	0	0	0.548765	0.094072	1	0.130738	0.048314	0.384022	0.087314	0.084237	0.0
6	1	0	0	0	0	0.589939	0.263537	0.972501	0.705793	0.10061	1	0.181402	0.385671	0.2
7	1	1	1	0	0	0.807692	0.048154	1	0.084615	0.161538	0.423077	0.038462	0.315385	0.2
8	3	1	3	0	0	0.892666	0.160095	1	0.274763	0.063765	0.814652	0.091036	0.151456	0.1
9	5	0	0	0	0	0.67702	0.16947	0.449973	0.283484	0.093472	1	0.116756	0.184678	0.1
10	10	1	10	0	0.001145	1	0.062602	0.272852	0.076581	0.027031	0.397883	0.035259	0.049178	0.0
11	2	0	0	0	0	0.701671	0.325652	0.593659	0.401201	0.083416	1	0.270312	0.134062	0.2
12	4	0	0	0	0	0.585976	0.201684	0.698887	0.327029	0.059685	1	0.153016	0.12643	0.1
13	1	0	0	0	0	0.810256	0.305128	1	0.412821	0.002564	0.958974	0.269231	0.010256	0.2
14	1	0	0	0	0	0.8742	0.347548	0.729211	0.663113	0.132196	1	0.289979	0.249467	0.2

Model 9	Test	Validity
Sensitivity	100%	100%
Specificity	100%	100%
Normal State	30/30 (100%)	37/37 (100%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 7048.018	8602.237	8654.395	1144.796	4260.403
0	28	1	28	0	0.117795	1	0.189136	0.00018	0.038846
1	4	0	0	0	0.44898	1	0.724911	0	0.518048
2	3	0	0	0	0.616286	0.893434	0.914925	0	0.472577
3	12	1	9	3	0.191145	1	0.325061	0	0.169693
4	7	0	1	1	0.214739	1	0.50704	0	0.340581
5	9	1	9	0	0.3496	1	0.389951	0	0.221401
6	4	0	0	0	0.745345	1	0.898562	0	0.634987
7	1	0	0	0	1	0.740741	0.618519	0	0.522222
8	1	1	1	0	0.646484	1	0.373047	0	0.307111
9	1	0	0	0	0.48337	0.846868	1	0	0.697436
10	2	0	0	0	0.515608	1	0.903216	0	0.728896
11	3	0	0	0	0.739768	1	0.882573	0	0.944444
12	1	1	1	0	0.513566	1	0.25969	0	0.108527
13	1	0	0	0	0.346457	1	0.602362	0	0.675197
14	1	0	0	0	0.933148	1	0.793872	0	0.465181

WO 2005/011474

PCT/US2004/024113

21

Model 10	Test	Validity
Sensitivity	100%	100%
Specificity	100%	100%
Normal State	30/30 (100%)	37/37 (100%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 7202.716	6004.417	7060.121	1001.654	1255.593	9387.113	4377.854	8605.678	8709.5
0	4	1	4	0	0.173188	0.074963	0.970492	0.003208	0.042568	0.29361	0.14722	0.858894	0.371
1	2	0	0	0	0.319725	0.176894	0.393018	0	0.164671	0.825969	0.379272	0.917131	0.667
2	10	1	10	0	0.199442	0.082052	0.660658	0	0.055131	0.403149	0.151314	1	0.459
3	2	0	1	1	0.361857	0.113665	1	0	0.121266	0.562191	0.202878	0.70216	0.829
4	2	1	2	0	0.213106	0.072628	0.578857	0	0.050348	0.562743	0.155164	1	0.5021
5	1	1	1	0	0.284091	0.113636	0.940341	0	0.150568	0.605114	0.207386	1	0.4716
6	3	1	3	0	0.263962	0.121837	0.831316	0	0.080509	0.411379	0.183044	1	0.6016
7	7	1	5	2	0.235242	0.08713	0.876821	0	0.082517	0.506915	0.140705	1	0.866
8	2	1	2	0	0.227143	0.128887	1	0	0.061198	0.421919	0.159605	0.619174	0.385
9	2	0	0	0	0.280298	0.087375	0.746558	0	0.068565	0.418376	0.128141	0.52401	
10	1	0	0	0	0.564168	0.180432	0.791814	0	0.15758	0.302414	0.123253	0.472681	
11	1	1	1	0	0.383361	0.168028	0.71615	0	0.174551	0.597064	0.17292	0.882055	
12	2	1	2	0	0.254143	0.094635	1	0	0.04466	0.198106	0.105068	0.463184	0.430
13	2	1	2	0	0.484788	0.101004	0.847496	0	0.088878	0.386489	0.190463	1	0.822
14	1	1	1	0	0.303093	0.053808	0.485979	0	0.083505	0.313402	0.130928	1	0.904
15	1	1	1	0	0.237782	0.167832	1	0	0.125874	0.454545	0.202797	0.825175	0.573
16	2	0	0	0	0.335049	0.15409	0.489544	0	0.070396	0.522135	0.262555	0.833444	0.971
17	2	1	2	0	0.359659	0.068265	1	0	0.105538	0.508054	0.173701	0.930654	0.874
18	2	0	0	0	0.243242	0.067837	0.335432	0	0.106513	0.341438	0.109465	0.518447	
19	8	1	8	0	0.123575	0.048128	0.311115	0	0.045892	0.286063	0.113572	1	0.382
20	2	0	0	0	0.211598	0.059312	0.548008	0	0.113593	0.450127	0.132826	0.790771	

WO 2005/011474

PCT/US2004/024113

21	4	0	0	0	0.329776	0.110944	0.509651	0	0.132027	0.484959	0.19387	0.587533						
22	1	0	0	0	0.253837	0.126328	0.291617	0	0.11098	0.5183	0.20307	1	0.9185					
23	1	0	0	0	0.601351	0.344595	0.783514	0	0.096847	0.86036	0.481882	0.878378						
24	1	0	0	0	0.329101	0.116402	0.569312	0	0.076191	0.274074	0.111111	0.394709						
25	2	0	0	0	0.453481	0.170665	0.800839	0	0.116823	0.618036	0.254696	0.552077						
26	3	1	3	0	0.119065	0.10091	0.491402	0	0.082836	0.204372	0.145723	1	0.2959					
27	1	0	0	0	0.178475	0.119283	0.300448	0	0.101345	0.917489	0.220828	0.673543						
28	1	0	0	0	0.554658	0.297571	0.870445	0	0.109312	0.534413	0.317814	0.720648						
29	1	1	1	0	0.083564	0.030732	0.097721	0	0.02797	0.11882	0.058356	1	0.3080					
30	1	0	0	0	0.457023	0.180294	0.57652	0	0.125786	0.574423	0.400419	0.698113						
31	1	0	0	0	0.678325	0.276371	0.736287	0	0.187764	0.601266	0.398734	0.879747						
32	1	1	1	0	0.169982	0.060579	0.289331	0	0.063291	0.352622	0.136528	1	0.6084					

WO 2009011492

PCT/US2008/070413

Model 11	Test	Validity
Sensitivity	100%	100%
Specificity	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error										
0	5	1	5	0	0.149439	1	0.249501	0	0.340138	0.141393	0.173882	0.219088	0.066197	0.2212
1	1	0	0	0	0.009091	0.94697	1	0	0.911616	0.578283	0.626263	0.348485	0.199495	0.3888
2	2	1	2	0	0.129668	0.75439	0.351176	0	0.304239	0.211129	0.215195	1	0.061103	0.1517
3	1	1	1	0	0.023943	0.454698	0.096057	0	0.162752	0.097735	0.097315	1	0.020554	0.0645
4	3	0	0	0	0.523752	0.968483	0.686268	0	0.990886	0.326104	0.594814	0.382	0.148411	0.4047
5	6	1	6	0	0.192401	1	0.497082	0	0.64152	0.256213	0.315258	0.32065	0.122937	0.3918
6	1	1	1	0	0.164719	1	0.943894	0	0.574257	0.339934	0.277228	0.749175	0.052805	0.3663
7	2	1	2	0	0.212839	1	0.329502	0	0.556667	0.202068	0.235864	0.628961	0.031436	0.1278
8	4	1	2	2	0.22784	1	0.410498	0	0.725683	0.218632	0.324713	0.331147	0.089938	0.2191
9	3	1	3	0	0.151335	0.945748	0.506252	0	0.438843	0.294054	0.316824	0.965705	0.028208	0.2972
10	1	0	0	0	0.280282	1	0.427657	0.134443	0.496799	0.276569	0.385403	0.18822	0	0.2132
11	1	1	1	0	0.324855	1	0.244728	0	0.447257	0.35865	0.329114	0.227848	0.046414	0.4218
12	2	0	0	0	0.03223	0.831889	0.951855	0	0.99322	0.441819	0.734281	0.576025	0.165179	0.2786
13	1	1	1	0	0.186281	1	0.785124	0	0.444215	0.289256	0.340909	0.21281	0.115702	0.3863
14	4	1	4	0	0.024548	1	0.686663	0	0.687229	0.222129	0.419095	0.487583	0.148942	0.3785
15	1	1	1	0	0.053571	1	0.805357	0	0.830357	0.348214	0.648214	0.594643	0.201786	0.5327
16	2	1	2	0	0.239768	0.991269	0.374159	0	0.739857	0.272116	0.351161	0.985558	0.135604	0.2246
17	2	1	2	0	0.457544	0.81331	0.338888	0	0.561209	0.189787	0.31758	0.987784	0.059326	0.1358
18	1	1	1	0	0.284549	0.678112	1	0	0.274678	0.206009	0.27897	0.077253	0.128755	0.2833
19	1	0	0	0	0.257671	1	0.219178	0	0.880626	0.223092	0.315068	0.260274	0.058708	0.1645
20	1	1	1	0	0.450685	1	0.676712	0	0.471233	0.30411	0.350685	0.745205	0.210959	0.2528

WO 2009011492

PCT/US2008/070413

21	2	0	0	0	0.325044	0.9441	0.869557		0	1	0.337525	0.940181	0.381157	0.190878	0.3377
22	7	1	6	1	0.38207	1	0.306461		0	0.551426	0.178093	0.223762	0.212152	0.090816	0.1941
23	3	0	0	0	0.155007	1	0.573712		0	0.801758	0.254962	0.719139	0.318285	0.190355	0.3333
24	1	0	0	0	0.594298	1	0.563596		0	0.855263	0.532895	0.434211	0.475877	0.236842	0.3442
25	2	1	2	0	0.184238	1	0.55516	0.083929	0.433112	0.278979	0.2783	0.612885	0.009195	0.2719	
26	2	0	0	0	0.029325	0.812591	0.702828		0	1	0.325657	0.802751	0.403117	0.171496	0.3254
27	1	0	0	0	0.2737	1	0.227829		0	0.640673	0.30581	0.440367	0.366972	0.114679	0.2461
28	1	1	1	0	0.344562	1	0.547988		0	0.668731	0.390093	0.340557	0.804954	0.065016	0.3126
29	1	0	0	0	0.357796	1	0.443425		0	1	0.357798	0.82263	0.477064	0.140673	0.339
30	1	0	0	0	0.040255	0.897996	0.464481		0	1	0.48816	0.741348	0.990393	0.127505	0.3387
31	2	0	0	0	0.236715	0.975111	0.493956		0	0.956469	0.375087	0.519177	0.279772	0.161407	0.3223
32	1	0	0	0	0.625316	0.920888	0.300633		0	0.893671	0.389241	0.512658	1	0.018987	0.5189
33	3	1	3	0	0.097097	1	0.183182		0	0.445501	0.125676	0.182921	0.453176	0.055861	0.1002
34	1	1	1	0	0.079783	1	0.144424		0	0.26891	0.096892	0.122486	0.025584	0.074954	0.0678
35	1	0	0	0	0.042891	0.884929	0.646919		0	1	0.400474	0.599526	0.575829	0.151659	0.2725
36	1	1	1	0	0.174923	1	0.517028		0	0.377709	0.227554	0.309598	0.28483	0.085139	0.3297
37	1	1	1	0	0.434043	1	0.612766		0	0.559574	0.282979	0.353191	0.653192	0.131915	0.3914
38	1	0	0	0	0.203571	0.748016	0.565476		0	1	0.328365	0.414683	0.337302	0.188492	0.1984

Model 12	Test	Validity
Sensitivity	100%	100%
Specificity	100%	95%
Normal State	30/30 (100%)	35/37 (95%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	#V2					
					88852	8709.548	7065.771	1132.049	8605.678	
0	6	1	6	0	0.227355	0.285099	0.294878	0	1	
1	2	0	1	1	0.579419	0.996676	0.249831	0	0.904368	
2	5	1	5	0	0.286212	0.48104	0.337354	0	1	
3	2	0	0	0	0.839555	1	0.545907	0	0.694336	
4	2	1	2	0	0.444594	0.494724	0.255931	0	1	
5	7	1	7	0	0.328116	0.404957	0.471929	0	1	
6	3	1	3	0	0.420975	0.599319	0.470769	0	1	
7	6	1	4	2	0.61664	0.902203	0.355835	0	1	
8	3	0	0	0	0.653035	0.84379	0.223522	0	1	
9	1	1	1	0	0.545	0.545	0.9675	0	1	
10	4	0	0	0	0.430854	1	0.405585	0	0.471429	
11	1	0	0	0	0.155009	1	0.448905	0	0.215501	
12	11	1	11	0	0.261647	0.357539	0.14853	0	1	
13	1	1	1	0	0.650505	1	0.39596	0	0.977778	
14	1	1	1	0	0.313343	0.812564	1	0	0.830565	
15	2	1	2	0	0.640593	0.804083	0.442778	0	1	
16	1	0	0	0	0.771379	1	0.319372	0	0.91274	
17	2	1	2	0	0.395313	0.746381	0.349265	0	1	
18	2	0	0	0	0.358251	1	0.141059	0	0.455628	
19	2	0	0	0	0.357038	1	0.251898	0	0.762878	
20	1	0	0	0	0.966006	1	0.68272	0	0.847026	

21	1	0	0	0	0.334625	1	0.31137	0	0.260982
22	1	1	1	0	0.376206	0.533762	1	0	0.951769
23	2	0	0	0	0.356085	1	0.272623	0	0.537859
24	2	0	0	0	0.579131	1	0.240333	0	0.640437
25	1	0	0	0	0.471058	1	0.660679	0	0.51497
26	1	0	0	0	0.56581	1	0.398458	0	0.62982
27	1	1	1	0	0.619256	0.833698	0.659584	0	1
28	1	0	0	0	0.782258	1	0.628032	0	0.846774
29	1	1	1	0	0.516	1	0.518	0	0.898
30	1	1	1	0	0.403558	0.594569	0.152622	0	1

Model 13	Test	Validity
Sensitivity	100%	100%
Specificity	100%	95%
Normal State	30/30 (100%)	35/37 (95%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 11098.07	6501.799	2087.371	8505.678	8888.674	7048.838	4262.107
0	8	1	6	0	0.053842	0.050308	0	1	0.277113	0.258017	0.126978
1	1	0	0	0	0.194368	0.016901	0	1	0.780282	0.24507	0.416901
2	1	0	0	0	0.230024	0.179177	0	1	0.890315	0.736077	0.493947
3	8	1	6	2	0.047783	0.03069	0.000757	1	0.473931	0.24506	0.11983
4	10	1	9	1	0.074636	0.064462	0	1	0.43221	0.343755	0.20137
5	8	1	7	1	0.094925	0.130769	0	1	0.971894	0.378017	0.273367
6	1	1	1	0	0.059567	0.032491	0	1	0.644404	0.355596	0.034296
7	1	0	0	0	0.230797	0.139693	0	1	0.830324	0.189319	0.459966
8	1	1	1	0	0.205333	0.056	0	1	0.514667	0.784667	0.122667
9	1	0	0	0	0.108929	0.123214	0	0.921429	1	0.883929	0.457143
10	1	0	0	0	0.068063	0.408377	0	0.832461	0.997382	1	0.505238
11	12	1	12	0	0.0376	0.018129	0.005735	1	0.282722	0.108974	0.075537
12	1	1	1	0	0.066466	0.115332	0	0.82768	0.499322	1	0.238806
13	1	1	1	0	0	0.082474	0.195876	1	0.402062	0.237113	0.154639
14	1	0	0	0	0.12328	0.280316	0	1	0.852883	0.274354	0.310139
15	2	0	0	0	0.043452	0.088573	0	1	0.935869	0.380821	0.614702
16	1	0	0	0	0.124457	0.859334	0	1	0.609262	0.357453	0.442284
17	1	0	0	0	0.192394	0.127517	0	0.876957	1	0.438479	0.628635
18	1	0	0	0	0.091245	0.165228	0	1	0.641184	0.181258	0.282367
19	1	0	0	0	0	0.313726	0.124183	0.85098	1	0.650327	0.441176
20	1	1	1	0	0.153302	0.179245	0	1	0.415094	0.566038	0.235849

21	1	0	0	0	0.128713	0.165842	0	0.759901	1	0.675743	0.537129	
22	2	0	0	0	0.194312	0.20655	0	0.94264	1	0.528225	0.430212	
23	1	1	1	0	0.2125	0.2	0	0.905	0.47	1	0.19	
24	1	0	0	0	0.270089	0.084821	0	0.841518	1	0.870536	0.546875	
25	1	0	0	0	0.134441	0.128399	0	0.960383	1	0.311178	0.303625	
26	1	0	0	0	0.397436	0.339744	0	0.858974	1	0.903846	0.490385	
27	1	0	0	0	0	0.257908	0	0.924574	1	0.491484	0.593674	
28	1	0	0	0	0.29085	0.382745	0	1	0.973856	0.990196	0.470588	
29	1	0	0	0	0	0.147287	0.036176	0.976744	1	0.50646	0.423773	
30	1	0	0	0	0.047222	0.175	0	0.76	1	0.497222	0.480556	
31	1	1	1	0	0.18996	0.278656	0	1	0.733202	0.743083	0.320158	
32	1	0	0	0	0.061404	0.285088	0	0.813596	1	0.698684	0.33114	
33	1	1	1	0	0.090909	0.130165	0	1	0.733471	0.607438	0.209678	

WO 2005011474

PCT/US2004/024113

32

Model 14	Test	Validity
Sensitivity	100%	100%
Specificity	100%	97%
Normal State	30/30 (100%)	36/37 (97%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 4162.719	8588.487	6709.548	8664.385	1319.958	8605.678	2280.256	7060.121
0	3	1	3	1	0	0.095692	0.344856	0.319228	0.242556	0.007524	0.969059	0.009948	0.959932
1	1	0	0		0	0.488175	0.68894	1	0.626728	0	0.880184	0.004608	0.31106
2	5	1	5		0	0.117272	0.439504	0.401233	0.30528	0	1	0.039692	0.853983
3	6	1	6		0	0.085015	0.498557	0.325581	0.28407	0.00115	1	0.014817	0.410254
4	1	0	0		0	0.153971	0.58671	0.85624	0.664506	0	0.862885	0.006483	1
5	1	1	1		0	0.109524	0.591667	0.504762	0.657143	0	1	0.105952	0.55
6	3	1	3		0	0.127988	0.493341	0.417644	0.3649	0.002772	0.984158	0.050381	0.925263
7	2	1	2		0	0.207404	0.724887	0.602078	0.532475	0	1	0.037808	0.814917
8	7	1	5		2	0.176899	0.715138	0.612647	0.551972	0.005477	0.998362	0.018468	0.650556
9	1	0	0		0	0.697282	0.824477	0.827697	0.68599	0	1	0.119163	0.310789
10	1	1	1		0	0.108787	0.428778	0.381227	0.403068	0	0.559275	0.026499	1
11	2	1	2		0	0.106972	0.628005	0.453237	0.363568	0.005034	1	0.030471	0.406813
12	3	0	0		0	0.152024	0.439381	1	0.428457	0.005728	0.479396	0.0065	0.730046
13	1	1	1		0	0.109208	0.304069	0.432548	0.248253	0	0.441114	0.068523	1
14	2	1	1		1	0.253559	0.857705	0.891482	0.692764	0.013306	1	0.006591	0.449839
15	1	1	1		0	0.242188	0.335938	0.523438	0.328125	0	0.804688	0.226562	1
16	1	0	0		0	0.225275	0.807692	1	0.723443	0.021978	0.908425	0	0.448718
17	1	1	1		0	0.182909	0.587706	0.890555	0.605697	0.043478	0.928036	0	1
18	1	1	1		0	0.14269	0.621053	0.788421	0.492398	0.014035	1	0	0.817544
19	2	0	0		0	0.172991	0.469996	1	0.484749	0.004406	0.484017	0	0.287822
20	5	1	5		0	0.062151	0.474033	0.407928	0.324867	0	1	0.013184	0.257672

WO 2005011474

PCT/US2004/024113

30

21	2	0	0	0	0.16018	0.506442	1	0.439991	0.008219	0.7738	0	0.511529	
22	3	1	3	0	0.153858	0.656383	0.450659	0.432758	0.004074	1	0.033124	0.717648	
23	1	1	1	0	0.2021	0.645669	0.703412	0.671916	0.026247	1	0	0.655843	
24	4	0	0	0	0.2007	0.575951	1	0.530549	0	0.522931	0.024103	0.458878	
25	1	0	0	0	0.209799	0.757539	0.913317	0.604271	0	1	0.035178	0.246231	
26	2	0	0	0	0.387106	0.8472	1	0.835188	0	0.850562	0.070583	0.702616	
27	1	1	1	0	0.164818	0.438986	0.29794	0.282092	0	0.729002	0.041204	1	
28	1	0	0	0	0.132353	0.438914	1	0.535973	0	0.352941	0.001131	0.539593	
29	1	1	1	0	0.123829	0.300728	0.240375	0.207078	0.009385	0.37565	0	1	
30	2	0	0	0	0.222129	0.625428	1	0.575785	0	0.504059	0.049331	0.778349	
31	1	0	0	0	0.101695	0.62343	1	0.57328	0	0.637089	0.041874	0.222333	
32	1	0	0	0	0.232258	0.873118	1	0.812903	0	0.703226	0.124731	0.862366	
33	2	1	2	0	0.132722	0.535895	0.83435	0.388105	0.008025	1	0.01613	0.569726	
34	1	1	1	0	0.035639	0.539873	0.292872	0.246295	0	1	0.000706	0.077982	
35	1	0	0	0	0.306122	0.716837	1	0.665816	0	0.632653	0.030612	0.484694	
36	1	1	1	0	0.210428	0.724395	0.787709	0.581006	0	0.929236	0.130354	1	
37	1	1	1	0	0.154391	0.627479	0.787535	0.505099	0.031162	0.715297	0	1	
38	1	1	1	0	0.070746	0.628195	0.586042	0.378585	0	1	0.016296	0.248566	

Model 15	Test	Validity
Sensitivity	100%	100%
Specificity	100%	100%
Normal State	39/39 (100%)	37/37 (100%)
Ovarian Cancer State	57/57 (100%)	40/40 (100%)

Node	Count	State	StateSum	Error	m/z 9870.936	2374.244	1278.881	7960.121	4292.9	8706.065	8605.678
0	33	1	33	1	0.120039	0.024623	0.01125	0.949945	0.171834	0.527519	0.872924
1	23	1	16		7 0.141653	0.02381	0.020585	0.528664	0.162589	0.626018	0.999723
2	7	0	2		2 0.188489	0	0.153321	0.882675	0.152271	0.953348	0.714632
3	16	0	1		1 0.144659	0	0.131107	0.595845	0.178005	1	0.741938
4	3	1	3		0 0.056997	0	0.043224	1	0.088753	0.359943	0.468551
5	1	1	1		0 0.04065	0	0.000353	0.076352	0.138211	0.276423	1
6	1	0	0		0 0.358639	0.146597	0	0.337696	0.397906	1	0.884293

What is claimed is:

1. A model usable in determining whether a biological sample taken from a subject indicates that the subject has ovarian cancer, comprising:

a vector space having at least three dimensions; and

at least one diagnostic cluster defined in said vector space, said diagnostic cluster corresponding to one of a diseased cluster and a healthy cluster,

said vector space having a first dimension that corresponds to a first mass to charge ratio value from a mass spectrum, said first mass to charge ratio being about 7060, said vector space having a second dimension that corresponds to a second mass to charge ratio value from a mass spectrum, said second mass to charge ratio being about 8605, and said vector space having a third dimension that corresponds to a third mass to charge ratio value from a mass spectrum, said third mass to charge ratio being about 8706.

2. The model of claim 1, wherein the vector space has at least four dimensions, said vector space having a fourth dimension that corresponds to a fourth mass to charge ratio value from a mass spectrum, said fourth mass to charge ratio being about 6548.

3. A model usable in determining whether a biological sample taken from a subject indicates that the subject has ovarian cancer, comprising:

a vector space having at least three dimensions; and

at least one diagnostic cluster defined in said vector space, said diagnostic cluster corresponding to one of a diseased cluster and a healthy cluster,

said vector space having a first dimension that corresponds to a first mass to charge ratio value from a mass spectrum, said first mass to charge ratio being about 9807, said vector space having a second dimension that corresponds to a second mass to charge ratio value from a mass spectrum, said second mass to charge ratio being about 2374, and said vector space having a third dimension that corresponds to a third mass to charge ratio value from a mass spectrum, said third mass to charge ratio being about 1276.

4. The model of claim 3, wherein the vector space has at least four dimensions, said vector space having a fourth dimension that corresponds to a fourth mass to charge ratio value from a mass spectrum, said fourth mass to charge ratio being about 4292.

5. A method of determining whether a biological sample taken from a subject indicates that the subject has ovarian cancer by analyzing the biological sample to obtain a data stream that describes the biological sample, comprising:

a. abstracting the data stream to produce a sample vector that characterizes the data stream in a predetermined vector space containing a diagnostic cluster, the diagnostic cluster being an ovarian cancer cluster, the ovarian cancer cluster corresponding to the presence of ovarian cancer;

b. determining whether the sample vector rests within the ovarian cancer cluster; and

c. if the sample vector rests within the ovarian cancer cluster, identifying the biological sample as being taken from a subject that has ovarian cancer.

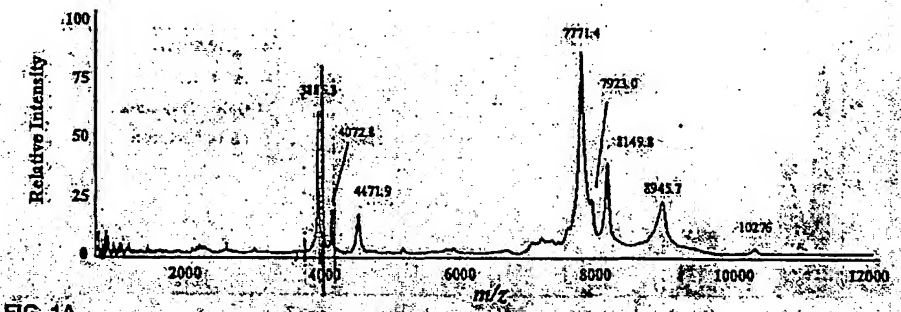


FIG. 1A

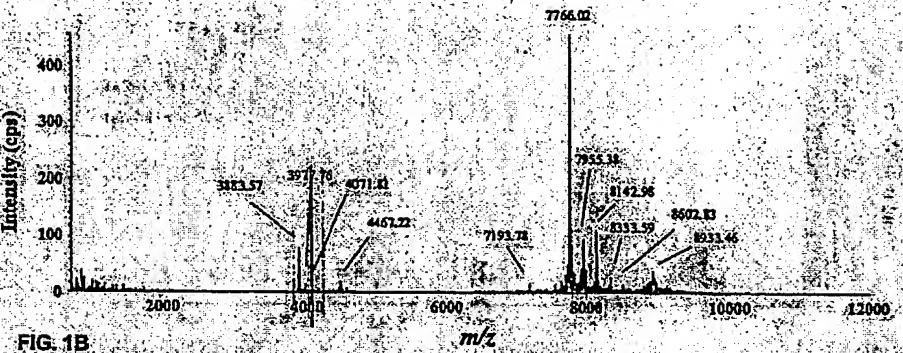


FIG. 1B

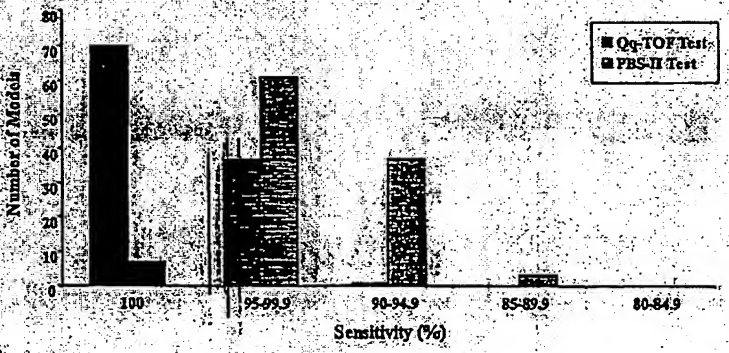


FIG. 2A

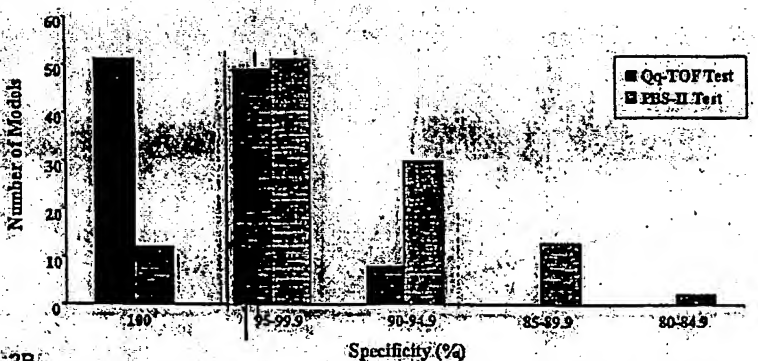


FIG. 2B

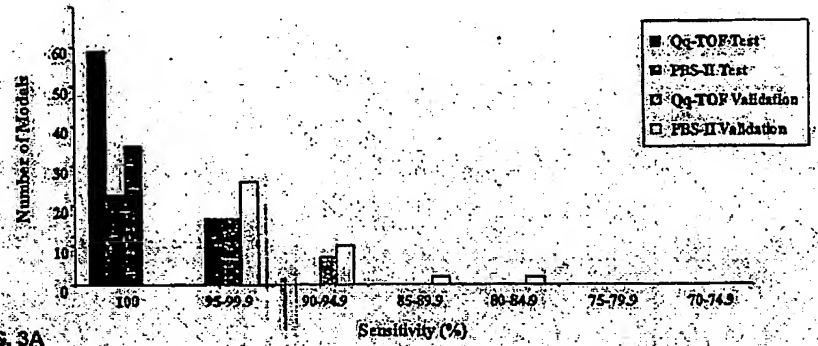


FIG. 3A

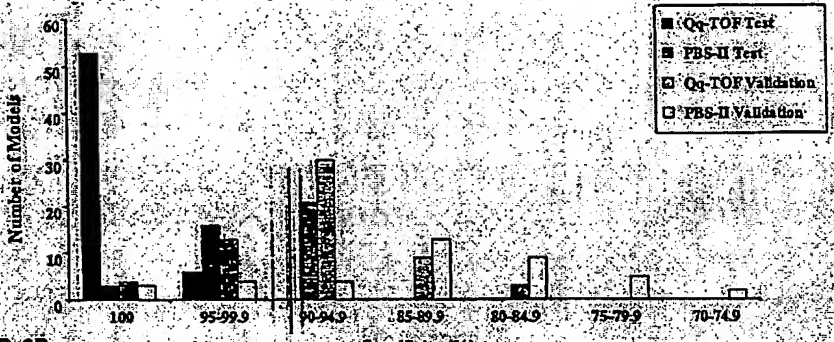


FIG. 3B

WO 2005/011474 PCT/US2004/024413

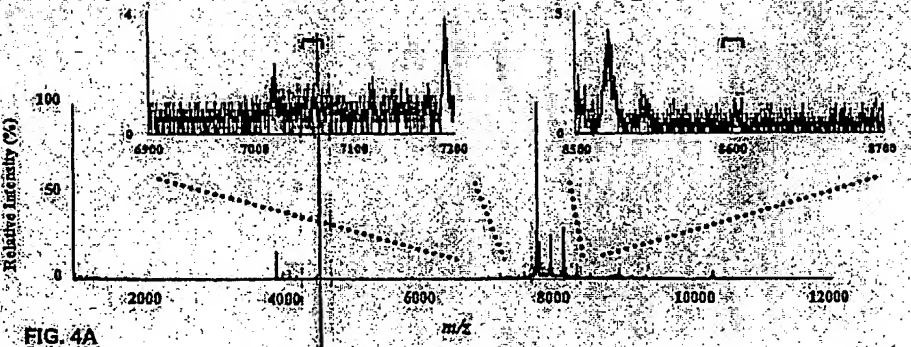


FIG. 4A

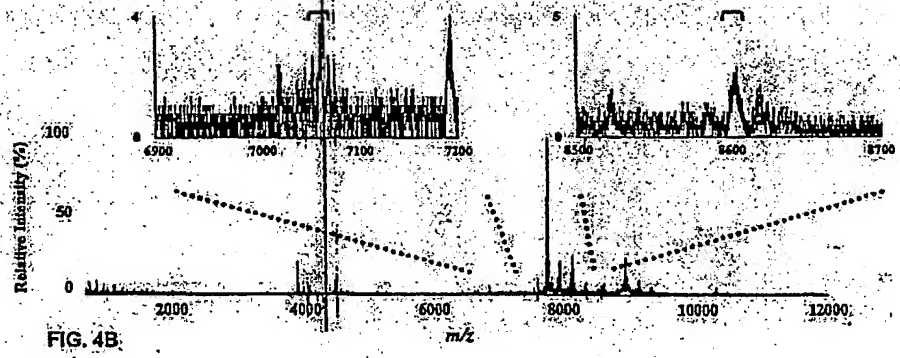


FIG. 4B